(51) International Patent Classification 6 :		(11) International Publication Number:	WO 97/20197
G01N	A2		
		(43) International Publication Date:	5 June 1997 (05.06.97)

(21) International Application Number: PCT/GB96/02959

(22) International Filing Date: 29 November 1996 (29.11.96)

(30) Priority Data:
9524381.2   29 November 1995 (29.11.95)   GB

(71) Applicant (for all designated States except US): THE ANTHONY NOLAN BONE MARROW TRUST [GB/GB]; The Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG (GB).

(72) Inventors; and
(75) Inventors/Applicants (for US only): ARGUELLO, Rafael [MX/GB]; The Anthony Nolan Bone Marrow Trust, The Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG (GB). AVAKIAN, Hovanes [GB/GB]; The Anthony Nolan Bone Marrow Trust, The Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG (GB). MADRIGAL, Alejandro [MX/GB]; The Anthony Nolan Bone Marrow Trust, The Royal Free Hospital, Pond Street, Hampstead, London NW3 2QG (GB).

(74) Agent: CAMPBELL, Patrick, John, Henry; J. A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5LX (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

(54) Title: METHOD FOR IDENTIFYING AN UNKNOWN ALLELE

(57) Abstract

The invention provides a method for identifying an unknown allele of a polyallelic gene, which method comprises (i) contacting the unknown allele with a panel of probes, each of which recognises a sequence motif that is present in some alleles of the polyallelic gene but not in others; (ii) observing which probes recognise the unknown allele so as to obtain a fingerprint of the unknown allele; and (iii) comparing the fingerprint with fingerprints of known alleles. The use of a panel of probes which each recognises a different motif allows identification of which motifs are present in the unknown allele. The alleles of the polyallelic gene each have a unique combination of motifs and so identification of this combination (or "fingerprint") leads to identification of the unknown allele.

METHOD FOR IDENTIFYING AN UNKNOWN ALLELE

The invention relates to a method and a kit for identifying an unknown allele of a polyallelic gene.

5

1. Background to the invention

1.1 General Introduction

10 Many genes exist as multiple alleles which differ from each other by small differences in sequence. It is sometimes desirable to identify an unknown allele of a polyallelic gene. For example, such identification is often necessary to match the alleles of the human leucocyte antigen (HLA) 15 genes in a prospective donor and a prospective recipient in a tissue or organ transplant operation; if the donor and recipient have the same HLA alleles, the probability of the recipient rejecting the donor's tissue is greatly reduced.

20 However, it can be a difficult task to identify precisely an unknown allele of a polyallelic gene because two alleles can differ from each other by as little as one nucleotide. The difficulties are increased in genes which have a very large number of different alleles, such as the major 25 histocompatibility complex (MHC) genes (e.g. the HLA class I genes which have 222 known alleles).

Up to date the most favourable bone marrow transplant (BMT) and kidney transplant results have been obtained using 30 sibling donors who are genotypically HLA-identical to the recipient but such donors are available for only about 30% of patients[1-5]. BMT using unrelated donors can be successful, but these transplants have higher rates of graft failure, increased incidence and severity of Graft 35 versus Host Disease and more frequent complications related to delayed or inadequate immune reconstitution [4].

- 2 -

New molecular biological methods for detection of genetic
polymorphism currently provide an opportunity to improve
matching of unrelated donors as well as a research tool to
investigate the relationship between genetic disparity and
5   transplant complications. These molecular typing methods
include sequence-specific amplification, hybridisation with
oligonucleotide probes, heteroduplex analysis, single
strand conformation polymorphism analysis and direct
nucleotide sequencing.

10

Each of these molecular approaches has been used for
routine HLA class II typing [6], but a variety of reasons
related to the HLA class I gene structure has complicated
and made relatively unsuccessful their application to class
15   I typing. The reasons for these complications are the
extensive polymorphism of class I and the degree of
sequence homology between the A, B and C loci of class I.
In addition, sequence homology between class I classical
and non-classical genes and the reported 12 pseudo genes
20   can cause problems for specific locus amplification [7].

The low occurrence of "allele specific" sequences at
polymorphic sites is a feature of the HLA class I genes
that has limited the resolution of all current DNA typing
25   approaches. An "allele specific" sequence is a sequence
that is only present in one allele and can therefore be
used to distinguish the allele from other alleles. The
occurrence on more than one exon of the specific sites for
determining the allelic specificity causes additional
30   problems in the identification of individual alleles. As a
result, there is at present no single method of typing
which can identify all HLA class I alleles of high
resolution; see Table A below.

Table A

Comparison of some of the currently used DNA typing methods for HLA class I alleles

Method	Reference	HLA locus	Reagents	Resolution
SSO	Dale et al '96 Tissue Antigens 47:93-101	A	91 probes	High
SSO	Fernandez-Villa '95 Tissue Antigens 45:153-168	B	99	Medium-high
SSO	Levine/Yang '94 Tissue Antigens 44:174-183	Cw	64 probes	Medium-high
SSP	Bunce et al '95 Tissue Antigens 46:355-367	A,B,Cw	104 primer mixes	Low (≥ serology)
SSP	Burce/Welsh '94 Tissue Antigens 43:7-17	Cw	22 primer mixes	Low (=serology)
SSP	Krausa et al '93 Tissue Antigens 42:91-99	A9, 10, 19, 28	30 primer mixes	Medium
SSP (nested)	Krausa et al '95 Tissue Antigens 45:223-231	A2	15 primer mixes	High
PCR-RFLP	Tazari et al '95 ProcNatlAcadSci 92:8803-7	Cw (partial. 23)	3 primer pairs, 11 endonucleases	High
SSP-SEQUENCING	Blasczyk et al '95 Tissue Antigens 46:86-95	A (40 alleles)	27 primer pairs automatic sequencing with fluorescent tracers	High
SSP-SEQUENCING	Petersdorf/Hansen '95 Tissue Antigens 46:73-85	B (98 alleles)	5 primer pairs automatic sequencing with fluorescent tracers	Medium-high ''
URSTO	Arguello et al '96 ProcNatlAcadSci 9 3:10961-5	A, B, Cw (201 alleles)	3 primer pairs 40 probes	High

- 4 -

Methods for allele separation

1.2   Sequence specific primer amplification (PCR-SSP)

5   This method utilises both the group-specific and, when
present, allele-specific sequence sites in PCR primer
design.   The SSP design is based on the amplification
refractory mutation system (ARMS), in which a mismatch at
the 3' residue of the primer inhibits non-specific

10   amplification[8,9].

Although each SSP reaction may not individually provide
sufficient specificity to define an allele, the use of
combinations of sequence specific primers allows the

15   amplification of their common sequences to give the desired
specificity.

However, despite its high accuracy, PCR-SSP is only in some
cases more informative than serology. The reason for this

20   is the low occurrence of allele specific sequence motifs in
the exons and this limitation has stimulated a vast amount
of research into the identification of allele specific
motifs even in the intron sequences[10].   However, up to
date this approach has not contributed considerably to the

25   identification of more alleles.

Another limitation of this method is that it detects a
limited number of polymorphic sequences which are utilised
to predict the entire sequence.   If an unknown allele is

30   present in a particular sample this extrapolation may be
incorrect.

In addition, the successful use of the technique relies on
group specific amplification and therefore prior knowledge

35   of broad HLA specificity is needed.

1.3   Single strand conformation polymorphism (SSCP)

This technique is based on the electrophoretic mobility of
single stranded nucleic acids in a non-denaturing
polyacrylamide gel, which depends mainly on sequence-

5   related conformation[11-13].   The technique can be employed
for isolating single alleles which could then be used for
further manipulation and analysis such as direct
sequencing.   The pattern of bands obtained after
electrophoresis may be diagnostic for an allele[14,15].

10

The major disadvantage of SSCP is the tendency of DNA
single strand to adopt many conformational forms under the
same electrophoretic conditions resulting in the presence
of several bands from the same product; this makes the

15   identification more difficult.   In addition there is a high
degree of variation and inconsistency in the sensitivity of
this method for detecting mutations or allelic variations
and there is a physical limitation in the size of the DNA
fragment which is of the order of 200-400 base pairs (16).

20

1.4   Denaturing Gradient Gel Electrophoresis (DGGE) and
Temperature Gradient Gel Electrophoresis (TGGE) (17,18)

The underlying principle of both techniques is the

25   difference in the degree of melting between two alleles
(double stranded DNA) which results in a reduction of
mobility of the DNA fragments in polyacrylamide gels
containing a denaturing reagent (DGGE) or a temperature
gradient (TGGE).

30

Both techniques have been used frequently for screening
mutations in genetic systems with one or two variants.
They are only rarely used for the separation of alleles in
highly polymorphic systems such as HLA.

35

Both techniques require specific conditions for a
particular system under investigation and, in addition,

where two alleles share common sequence segments with low
melting points they may not always be differentiated. The
simultaneous melting of both alleles will produce very
similar retardations.

5

1.5. Cloning of DNA

This is the classical method of preparation of a single
sequence, i.e. the sequence derived from a single allele.
10   A variety of constructs has been used to introduce the
required DNA fragment into a plasmid and grow sufficient
copies for analysis. This method yields pure samples of
the analyte, but is time consuming to perform and several
clones are normally tested to ascertain the homogeneity of
15   the product.

Methods for the identification of alleles

1.6   Heteroduplex analysis
20

Fully matched DNA duplexes are more stable than those with
base mismatches. Instability of the duplex increases with
the number of nucleotide mismatches; these cause formation
of loops and bends in the linear DNA fragment which produce
25   an increasing "drag effect" in polyacrylamide gels which
retard the affected migrating bands[19-21].

Mismatched DNA hybrids (heteroduplex) may be formed at the
end of each PCR cycle between coamplified alleles from a
30   particular locus or loci due to primer cross reaction at
sites with similar sequences. During the annealing stage
of each cycle of the PCR, a proportion of sense strands of
each allele may anneal to anti-sense strands of different
alleles. The banding pattern obtained in PAGE analysis can
35   be useful for identifying the alleles involved in the
reaction[22-24].

- 7 -

Heteroduplex analysis is an approach that has been utilised to compare HLA genes of a particular donor and recipient. HLA genes are amplified, denatured (melted into single strands) and mixed together under conditions that promote
5   renaturation to form double stranded molecules. If the HLA genes of a donor and recipient are similar but not identical, heteroduplexes will form consisting of one strand of an allele of donor origin and a second strand from a different allele of recipient origin[25,26].   The
10   sensitivity of this method can be increased by adding DNA from an HLA allele that is not present in the donor or recipient.

The major advantage of heteroduplex analysis is that it is
15   relatively easy and inexpensive. Limitations of this approach include inability to detect certain HLA disparities, potential detection of irrelevant silent mutations and lack of specific information regarding the nature of the alleles involved.
20

Up to date this approach has been used for HLA class II typing with limited success. Its application to class I typing has not been successful.

25   1.7   Sequence specific oligonucleotide probes (PCR-SSO)

SSO typing involves amplification of HLA alleles from a particular locus followed by hybridisation with a panel of oligonucleotide probes to detect polymorphic sequences that
30   distinguish one allele or group of alleles from all others. In polymorphic systems a one step operation may not always differentiate all the known alleles; selected primers can be used to achieve amplification of individual alleles which are then identified by specific probes. This second
35   stage of oligotyping is often referred to as high resolution oligotyping[6].

- 8 -

The advantages of the PCR-SSO method are specificity, sensitivity, simplicity, reproducibility, and it is relatively inexpensive to operate and allows simultaneous processing of many samples. This approach has been applied
5 successfully, for example to typing of HLA class II alleles.

The major methodological drawback of this approach is that the complexity of the technique is directly related to the
10 number of alleles under investigation and the presence of two alleles in the heterozygous condition can complicate the identification process.

Published oligotyping methods could result in incorrect
15 interpretation of data if certain combinations of recently discovered alleles are present in a specimen. It is therefore necessary to update the reagents used in the identification step.

20 Several typing approaches for HLA-A and B based on PCR-SSO have been published; these typically require over 40 and 90 probes respectively [27,28]. The operation of these methods is time consuming and the resolution obtained is only moderate.
25

1.8 Nucleotide Sequencing

DNA templates for sequencing can be produced by a variety of methods, the most popular being the sequencing of cloned
30 genomic or cDNA fragments, or the direct sequencing of DNA fragments produced solely by PCR (as in 1.2 above). These templates represent a single sequence derived from one haplotype. Alleles from both haplotypes of a heterozygous sample may be co-amplified and sequenced together using
35 locus-specific PCR primer.

The recent availability of computer software, which allows

the user to align the derived sequence against established
sequence libraries, has facilitated the analysis and allele
assignments for heterozygous samples in which both
templates are sequenced at the same time[27]. The
5 effectiveness of this method depends on the amount and
frequency of ambiguous heterozygous combinations; for
example there are many HLA class II alleles that when
present together in one sample cannot be differentiated by
this method. The number of such ambiguous combinations of
10 allele sequences is even greater for HLA class I alleles.

Up to date two HLA class I typing approaches based on
direct sequencing have been published. Both require
serology information followed by allele specific PCR
15 amplification and then direct sequencing[14,30]. More recent
practice, however, is to amplify DNA fragments without
prior knowledge of the allele groups and to use locus
specific PCR amplification. Theoretically these approaches
should give the highest resolution, but they are beset by
20 ambiguous sequence combinations which cannot be resolved
satisfactorily and in practice these methods are expensive
and difficult to perform routinely.

2.   Analysis of the HLA class I polymorphism
25

Genetic recombination plays a key role in the generation of
HLA alleles. This is supported by pairwise comparison of
the nucleotide sequences. The most closely related pairs
of alleles usually differ by localised clusters of
30 substitutions for which both sequence motifs can be found
in other alleles. This pattern implicates interallelic
conversion or double recombination as the diversifying
mechanism[7]. Although the vast majority of such events
appear to involve recombination between alleles of the same
35 locus, there are several cases that involve recombination
between alleles of different loci[31].

In comparison to the many pairs of alleles that differ by localised clusters of substitutions, few pairs differ by point substitutions and of these only a handful differ by a substitution that has not been found in another allele.

5   Thus, it appears that the rate at which point mutations create new alleles is slower than the rate at which new mutations are subsequently recombined with existing mutations (Figure 1).

10   Comparison of allelic HLA class I sequences[32] reveals substitutions throughout the coding region. There is, however, a higher frequency of substitutions within exons 2 and 3 which encode the α1 and α2 domains of the HLA molecule. In comparing pairs of HLA-A, B and C alleles

15   only 2 pairs out of a total of 6,460 possible combinations can not be distinguished on the basis of nucleotide sequences in exons 2, 3 and 4. However, if the comparison is restricted to exons 2 and 3 this number only increases to 5 pairs of ambiguous sequences. By contrast, when

20   comparison is restricted to either exon 2 or exon 3 alone then the number of ambiguous pairs increases significantly (Table B). This observation is relevant to the design of DNA-based methods for class I typing because it shows that for practical purposes all alleles can be discriminated on

25   the basis of sequence analysis of exon 2 and 3. Although there is some polymorphism in exon 4 encoding the α3 domain, mostly in HLA-A alleles, incorporating the information from exon 4 into the above analysis does not significantly increase the number of pairs for which the

30   alleles can be discriminated.

TABLE B

Discriminating HLA-A, B, C alleles on the basis of partial sequences

	Number of pairs	Exon 2,3 or 4 differ.	Exon 2 or 3 differences	Exon 2 differ.	Exon 3 differ.	Exon 4 differ.
Pairs without						
HLA-AXHLA-A	1081	0	1	26	28	129
HLA-BXHLA-B	4851	2	3	117	89	880
HLA-CXHLA-C	528	0	1	12	7	61

P. Parham (7)

- 12 -

In the development of PCR-based methodologies for the
detection of alleles, one of the most important steps is
the identification of primer sequences unique for the
target gene which includes all polymorphic sites of
5   interest in the amplified fragment, which should also be
manageable in length. Typing of the polymorphic sites in
exons 2 and 3 would facilitate the identification of all
recognised alleles of HLA-A, B and C loci, with 5
exceptions, if suitable locus-specific amplification could
10 be achieved.

Specificity of the primers should ensure the effective
amplification of target gene fragments. In practice
however, trace amplification of competing, cross-
15 hybridising templates may also take place. In addition, due
to the shared polymorphic sequence motifs between class I
alleles of all three loci, non-specific coamplification of
the DNA fragments would hinder specific identification. In
practice, it would therefore be advantageous to use a
20 method that allows the separation of the desired product
from the undesirable PCR fragments.

Within exons 2 and 3 of the HLA-A, B and C genes there are
only a few locus specific sites which are located primarily
25 in the central region of each exon which would restrict the
amplification to incomplete exon fragments. As discussed
above, this would reduce the allele specific information
necessary for the identification of all allelic variants.

30 The two polymorphic exons are flanked by introns 1 and 3,
and separated by intron 2. Thus, the ideal location for
primer sites to amplify exons 2 and 3 together as one
fragment would be within introns 1 and 3.

35 Cereb and collaborators[33] have described primer sequences
located in the first and third introns which can be used
for locus-specific amplification of the entire exon 2 and 3

- 13 -

region of the HLA-A, B and C genes in one fragment. Their
data indicated that the primers used in that study were
effective in the amplification of HLA-A, B and C genes.
Furthermore, the amplification was truly locus-specific, as
5   assessed by hybridisation with locus-specific, group-
specific, and allele-specific oligonucleotide probes.

   3.   Summary of the Invention

10 The invention provides a method for identifying an unknown
allele of a polyallelic gene, which method comprises

          (i)   contacting the unknown allele with a panel of
                probes, each of which recognises a sequence motif
15   that is present in some alleles of the
                polyallelic gene but not in others;

          (ii) observing which probes recognise the unknown
                allele so as to obtain a fingerprint of the
20   unknown allele; and

          (iii) comparing the fingerprint with fingerprints of
                known alleles.

25 The invention also provides a kit for identifying an
unknown allele of a polyallelic gene, which kit comprises a
panel of probes, each of which probes recognises a sequence
motif that is present in some alleles of the polyallelic
gene but not in others.   (The same motifs may also occur in
30 other loci in linked gene complexes with similar
exon/intron structures.)   The kit preferably also comprises
a database which indicates which probes in the panel
recognise each allele of the polyallelic gene.

35 The use of a panel of probes which each recognises a
different motif allows identification of which motifs are
present in the unknown allele.   The alleles of the

polyallelic gene (and alleles of other genes in a linked complex) each have a unique combination of motifs and so identification of this combination (or "fingerprint") leads to identification of the unknown allele. Thus, the

5 invention allows identification of alleles of polyallelic genes, such as the HLA class I genes, which may not contain "allele specific" sequences (i.e. individual sequences which are unique to one particular allele).

10 **4.   Brief Description of the Drawings**

Figure 1 is a schematic representation of a polyallelic gene which has evolved by recombination events and to which the invention can be applied. See the Detailed Description

15 of the Invention for more details.

Figure 2A shows a schematic overview of an embodiment of the "Complementary Strands Analysis" (CSA) technique that can be used to purify an allele for use in the method of

20 the invention.

Figure 2B shows results of this CSA technique. In particular, Figure 2B shows an autoradiograph of the separation of HLA-A,B and Cw alleles from three

25 International Histocompatibility Workshop cell lines by PAGE. Individual bands are eluted from the gel and used for subsequent analysis. Each band is a purified product from a single allele.

30 **Figure 3** shows the hybridisation pattern of an URSTO probe (number 37 from Table 1) with HLA-A, B and C allele products. DNA from 15 IHW cell lines was processed by complementary strand analysis into allelic products. These were blotted on nylon membranes and was hybridised with

35 URSTO probes (here 37). After washing the dots were developed and chemiluminescence were captured by autoradiography. In every case the presence of a signal

corresponds with the hybridisation patterns described in
Table 2. The unique aspect of this method is that the
alleles of the three loci can be identified simultaneously
by the analysis of the 40 membranes. 1-15 cell lines, C
5   control DNA product.

Figure 4   shows an HLA class I analysis of four cell lines
with 12 URSTO probes.

10   HLA types of the cell lines:
     a. L0541265 A*0101, B*0801, Cw*0701;
     b. STIELIN A*0101, B*0801, Cw0701
     c. LBUF A*3001, B*1302, Cw0602
     d. BER A*0201, B*1302, Cw*0602
15

HLA-A, B and Cw alleles are blotted on the same membrane
each membrane is hybridised with one probe. Locus specific
amplification and allelic separation of the amplified
fragments were performed as described in the Example below.
20   DNA was applied to nylon membranes and these were
hybridised with URSTO probes, and after washing the
chemiluminescence was recorded by autoradiography.

     5.   Detailed Description of the Invention
25

The invention can be applied to any polyallelic gene system
in which there are motifs that are present in some alleles
but not in others. The invention is mainly applicable to
polyallelic systems that have evolved by recombination
30   events and/or by gene conversion in polygenic linked
complexes. Examples of genes to which the invention can be
applied are the mammalian MHC genes (e.g. the HLA class I
and class II genes), the T cell receptor genes in
mammals[36,37], TAP, LMP, ras, nonclassical HLA class I
35   genes, human complement factor genes C4 and C2, Bf in the
HLA complex, and genes located in mitochondrial DNA,
bacterial chromosomes and viral DNA.

- 16 -

Figure 1 illustrates a motif pattern that could have evolved from two alleles, each with four motifs (rectangles), of an ancestral gene. There are also allele specific sequences on each allele (diamond shape) which may

5   have evolved from point mutations. The allele specific sequences are targets for SSO and SSP type allele identification techniques, but a minimum of 13 probes would be needed. In the method of the invention, only four probes (indicated on the top of the Figure) would be

10 required to type the entire range. New coherent patterns would indicate new unknown alleles as shown at the foot of the Figure.

Identification of the alleles is by the presence or absence

15 of hybridisation (+ or - respectively) of the probes as shown in Figure 3. Lack of binding, i.e. "-" in this system, is an important signal for pattern formation. No pattern is repeated in this 16 allele system and therefore each of the 16 alleles can be unambiguously identified by

20 the invention.

The invention is particularly applicable to HLA class I genes. Comparison of HLA-A, B, C allelic sequences reveals a patchwork pattern in which an individual allele comprises

25 a unique combination of sequence motifs, each of which is shared with other alleles, and only a few alleles have a specific sequence that is not present in other alleles (see Arnett and Parham (1995) Tissue Antigens 45 217-257). Many authors agree that this characteristic of the HLA class I

30 genes has limited the resolution of all current DNA typing approaches. This feature itself has been exploited to facilitate the identification of all known class I alleles.

Comparison of the sequences of all known HLA class I

35 alleles has led to the realisation that certain sequence motifs with one or more base substitutions recur in the same position in a locus and also in the same position in

another locus.   Each allele contains a unique combination
of these motifs; this feature is universal in all
polyallelic genes that have evolved mainly through
recombination events and/or by gene conversion.   It is
5   therefore possible to identify the alleles of such genes by
a limited number of selected recurring motifs.

Analysis of these common motifs in the HLA class I complex
on human chromosome 6 has led to the conclusion that by
10 selecting a limited number of motifs it would be possible
to identify all known alleles of this system by unique
hybridisation patterns from this selected panel.   Table 1
gives examples of sense probe sequences that identify these
motifs.   The probes could equally well have the antisense
15 sequences.

In essence therefore this method differs from any other
hitherto described method in that it does not target allele
specific regions of the gene (cf SSO and SSP) but utilises
20 recurring motifs which in specific combinations are unique
for each allele.

A very large number of allele specific motif patterns can
be generated with probes.   The number of motif patterns
25 generated by these oligonucleotides are sufficient to
identify at least 201 HLA class I alleles.   The sequences
of 40 oligonucleotides are given in Table 1 and the
expected patterns shown in Table 2.   Table 3 shows the
location and distribution of the 40 probes in HLA class I
30 genes.

The selection of the target motifs for these probes ensures
that for a coherent pattern no two probes for the same
sequence location can hybridise with the product from a
35 single allele.   Incoherent patterns indicate an error in
the amplification or separation stages.

For unambiguous pattern identification it is usually necessary to analyse the alleles individually. The use of Complementary Strands Analysis (see below) is provided as a means of separating amplified alleles from each other.

5

Table 1

Nucleotide sequences and the sites of URSTO probes for the HLA class I genes

ID	Sequence	Location
2	GGG CCG GCC GCG GGG AGC	113–130
3	CTC ACA GAT TGA CCG AGT	282–299
4	CGG ATC GCG CTC CGC TAC	307–324
5	TAC CTG GAG GGC CTG TGC	547–564
6	CAG AGG ATG TAT GGC TGC	358–375
7	ACA CCC TCC AGA GGA TGT	350–367
8	CAG AGG ATG TTT GGC TGC	358–375
9	CGA CGT GGG GCC GGA CGG	375–392
10	CTC ACA TCA TCC AGA GGA	347–364
11	TGT ATG GCT GCG ACC TGG	365–382
12	CCA GCA GGA CGC TTA CGA	411–428
13	GTG CGT GGA CGG GCT CCG	561–578
14	GCG GAC ACG GCG GCT CAG	478–495
15	GGA GCA GTG GAG AGC CTA	531–548
16	GGA GCA GTT GAG AGC CTA	531–548
17	GTG CGT GGA GTG GCT CCG	561–578
18	GGA GCA GCT GAG AGC CTA	531–548
19	AGG GGC CGG AGT ATT GGG	236–253
20	GGC CCG ACG GGC GCC TCC	382–400
22	TCC GCG GGC ATA ACC AGT	401–418
23	ACC AGT TCG CCT ACG ACG	413–430
24	ATT GGG ACC GGA ACA CAC	248–265
25	TAC CTG GAG GGC ACG TGC	557–574
26	TGT ATG GCT GCG ACG TGG	365–382
28	GCC CAG TCA CAG ACT GAC	277–292
29	ACC GAG TGG ACC TGG GGA	293–310
30	CGG AAC CTG CGC GGC TAC	307–324
31	ATT TCT ACA CCT CCG TGT	92–109
32	GCC CGT GTG GCG GAG CAG	520–537
33	GAT CTC CAA GAC CAA CAC	267–284
34	TGA CCA GTC CGC CTA CGA	411–428
35	AAC ACA CAG ATC TAC AAG	259–276
36	CGC GGG CGC CGT GGG TGG	212–229
37	AGA TAC CTG GAG AAC GGG	580–597
40	TCT CAC ACC CTC CAC	346–360
41	ACC AAC ACA CAG ACT GAC C	277–295
42	GGC GCA GCA GCG GAG A	528–543
43	CAG GAC GCC TAC GAC GGC	415–432
44	GAG GAC CTG CGC TCC TGG	454–471
45	GAA GGA GAC GCT GCA GCG	597–614

ID, identification number.

TABLE 2

Allele specific hibridisation patterns

Allele												
A*0101	2	9	13	19	26	37	42	43	45			
A*0102	9	13	19	26	37	42	43	45				
A*0201	2	6	16	17	25	26	29	37	45			
A*0202	2	6	7	15	17	25	26	29	37	40	45	
A*0203	2	6	15	17	25	26	29	37	45			
A*0204	2	16	17	25	26	29	37	45				
A*0205	2	6	7	15	17	25	26	29	31	37	40	45
A*0206	2	6	16	17	25	26	29	31	37	45		
A*0207	2	16	17	25	29	37	45					
A*0210	2	8	16	17	25	29	31	37	45			
A*0211	2	3	6	16	17	25	26	29	37	45		
A*0212	2	6	17	25	26	29	37	45				
A*0213	2	6	17	25	26	29	37	45				
A*0214	2	6	7	16	17	25	26	29	31	37	40	45
A*0215N	2	16	17	25	29	37	45					
A*0216	2	6	16	17	26	29	37	45				
A*0217	2	16	17	25	29	37	40	45				
A*0301	2	6	17	19	26	28	29	37	43	45		
A*0302	2	17	19	26	28	29	37	43	45			
A*1101	2	9	17	19	26	28	29	31	37	43	45	
A*1102	9	17	19	26	28	29	31	37	43	45		
A*2301	2	4	13	16	19	25	32	37	40	45		
A*2402	2	4	13	19	25	37	40	45				

Allele																
A*2403	2	4	17	19	25	37	40	45								
A*2404	2	13	19	25	37	40	45									
A*2405	2	4	13	19	25	37	40	45								
A*2406	2	4	13	15	19	25	37	40	45							
A*2407	2	4	13	19	25	28	37	40	45							
A*2501	2	4	6	9	12	15	17	19	24	26	31	37	45			
A*2601	2	6	9	12	15	17	19	24	26	31	37	45				
A*2602	6	9	15	17	19	24	26	31	37	45						
A*2603	2	6	9	12	15	17	19	24	26	29	31	37	45			
A*2604	2	5	6	9	12	15	17	19	24	26	31	37	45			
A*2605	2	6	9	12	15	17	19	24	26	31	37	45				
A*2901	2	16	17	19	25	28	32	37	43	45						
A*2902	2	16	17	19	25	28	32	37	43	45						
A*3001	16	17	25	26	28	29	37	45								
A*3002	16	17	25	26	37	45										
A*3003	16	17	19	25	26	37	45									
A*3004	15	17	25	26	37	45										
A*3005	15	17	25	26	37	45										
A*31011	2	3	16	17	25	26	29	32	37	43	45					
A*31012	2	3	16	17	25	26	29	32	37	43	45					
A*3201	2	4	9	16	17	19	25	26	32	37	43	45				
A*3301	2	3	16	17	19	24	25	26	32	37	43	45				
A*3302	3	16	17	19	24	25	26	29	32	37	43	45				
A*3303	2	3	16	17	19	24	25	26	29	32	37	43	45			
A*3401	2	6	9	12	15	17	19	24	25	26	28	29	31	37	45	

A*3402	2	16	17	19	24	25	26	28	29	31	37	45		
A*3601	2	9	17	19	25	26	37	42	43	45				
A*4301	2	6	9	12	15	17	19	26	31	37	45			
A*6601	2	6	9	12	15	17	19	24	26	28	29	31	37	45
A*6602	2	6	9	12	15	17	19	24	26	28	29	31	37	45
A*68011	2	15	17	19	24	25	26	28	29	31	37	43	45	
A*68012	2	15	17	19	24	25	26	28	29	37	43	45		
A*6802	2	6	9	15	17	19	24	25	26	28	29	37	45	
A*6901	2	6	16	17	19	24	25	26	28	29	31	37	45	
A*7401	2	9	16	17	19	25	26	29	32	37	43	45		
A*8001	2	13	18	26	37	43	45							

HLA-B alleles

Allele																
B*0702	2	9	14	17	19	24	30	31	35	37	40	42	44			
B*0703	2	9	14	17	19	24	30	31	35	37	40	41	42	44		
B*0704	2	9	14	17	19	24	30	31	35	37	40	44				
B*0705	2	9	14	17	19	22	24	30	31	35	37	40	42	44		
B*0801	2	9	17	19	22	24	25	30	32	37	40	41	44			
B*0802	2	9	17	19	22	24	25	32	37	40	41	44				
B*1301	2	6	10	11	17	18	19	22	32	33	37	45				
B*1302	2	11	17	18	19	22	32	33	37	45						
B*1303	2	5	11	17	18	19	22	32	33	37	45					
B*1401	2	9	17	18	23	24	25	26	30	31	40	41	45			
B*1402	2	9	17	18	23	24	25	26	30	40	41	45				
B*1501	2	5	7	9	14	15	17	19	30	33	34	37	40	45		
B*1502	2	5	6	9	10	14	17	19	24	26	30	33	34	37		
B*1503	2	5	7	9	14	17	18	19	30	33	34	37	40	45		
B*1504	2	5	9	14	15	17	19	26	30	33	34	37	45			
B*1505	2	5	7	9	14	17	18	19	30	32	33	34	37	40	45	
B*1506	2	5	7	8	9	14	17	19	30	33	34	37	40	45		
B*1507	2	5	9	14	15	17	19	30	33	34	37	40	45			
B*1508	2	5	7	9	14	17	19	24	30	34	37	40	45			
B*1509	2	5	7	9	14	17	18	19	22	24	30	37	40	45		

B*1510	2	5	7	9	14	17	18	19	24	30	37	40	37	40	45
B*1511	2	5	7	9	14	15	17	19	24	30	34	35	37	40	45
B*1512	2	5	7	9	13	14	15	19	30	33	34	37	40	45	
B*1513	2	4	5	6	9	10	14	17	18	19	24	26	33	34	37
B*1514	2	5	7	9	14	15	19	30	33	34	37	40	45		
B*1515	2	5	7	9	14	15	17	19	24	30	33	34	37	40	
B*1516	2	4	5	6	11	14	17	18	19	34	37	45			
B*1517	2	4	5	7	9	14	17	18	19	37	40	43	45		
B*1518	2	5	7	9	14	17	18	19	24	30	34	37	40	45	
B*1519	2	5	7	9	13	14	15	19	30	33	34	37	40	45	
B*1520	2	5	6	10	11	17	18	19	20	30	32	33	34	37	
B*1521	2	5	6	9	10	14	17	18	19	24	26	30	34	37	45
B*1522	2	5	7	9	14	15	17	19	24	30	34	37	40	45	
B*1523	2	4	5	7	9	14	17	18	19	24	34	37	40	45	
B*1524	2	4	5	7	9	14	15	17	19	33	34	37	40	45	
B*1525	2	5	6	9	10	14	17	18	19	26	30	33	34	37	
B*1526N	2	5	7	9	14	15	17	19	30	33	34	37	40	45	
B*1528	2	5	7	9	14	15	17	19	30	33	34	37	40	45	
B*1529	2	5	7	9	14	17	18	19	24	30	34	37	40	45	
B*1801	2	7	9	17	18	19	24	25	30	32	33	34	40	45	
B*1802	2	9	17	18	19	25	26	30	32	33	34	40	45		
B*2702	2	4	9	14	17	18	19	26	32	37	40	43	45		
B*2703	2	9	14	17	18	26	32	37	40	43	45				
B*27052	2	9	14	17	18	19	26	32	37	40	43	45			
B*27053	2	9	17	18	19	26	32	37	40	43	45				

B*2706	2	9	14	17	18	19	26	37	40	45						
B*2707	2	9	14	17	18	19	22	32	37	40	44	45				
B*2708	2	9	14	17	18	19	26	30	32	37	40	43	45			
B*2709	2	9	14	17	18	19	26	32	37	40	45					
B*3501	2	5	6	10	11	17	18	19	20	24	32	34	37			
B*3502	2	5	6	10	11	17	18	19	22	24	30	32	37	45		
B*3503	2	5	6	10	11	17	18	19	20	23	24	30	32	37	45	
B*3504	2	5	6	10	11	17	18	19	20	22	24	30	32	37		
B*3505	2	5	17	18	19	20	24	30	32	34	37	40	45			
B*3506	2	5	6	10	11	17	18	19	20	22	23	24	30	32	37	
B*3507	5	6	10	11	17	18	19	20	24	30	32	34	37	45		
B*3508	2	5	6	10	11	17	18	19	20	24	30	32	34	37	42	45
B*3509	2	5	6	10	11	17	18	19	20	22	24	30	32	37	44	45
B*3510	2	5	6	10	11	17	18	19	20	30	32	34	37	45		
B*3511	2	5	6	10	11	17	18	19	20	24	30	34	37	45		
B*3512	2	5	9	10	17	18	19	22	24	30	32	37	45			
B*3513	2	5	6	10	11	17	18	19	20	23	30	32	37			
B*3701	2	9	17	19	23	25	32	33	37	45						
B*3801	2	4	7	9	17	22	23	24	25	31	32	37	40	45		
B*3802	2	7	9	17	22	23	24	25	31	32	37	40	45			
B*39011	2	7	9	17	22	23	24	25	30	31	32	37	40	41	45	
B*39013	2	7	9	17	22	23	24	25	30	31	32	37	40	41	45	
B*39021	2	7	9	17	22	23	24	30	31	32	33	37	40	41	45	
B*39022	2	7	9	17	19	22	23	25	30	31	32	33	37	40	41	45
B*3903	2	9	17	22	23	24	25	30	31	32	37	40	41	45		

Allele															
B*3904	2	7	9	17	22	23	24	25	30	32	37	40	41	45	
B*3905	2	7	9	17	22	23	24	25	30	31	32	37	40	40	45
B*39061	2		9	17	22	23	24	25	30	31	32	37	41	45	
B*39062	2		9	17	22	23	24	26	30	31	32	37	41	45	
B*3907	2	7	9	17	24	25	30	31	32	34	37	40	45		
B*3908	2	7	9	17	22	23	25	30	31	32	33	37	40	42	45
B*40011	2	7	9	14	17	18	19	22	30	32	33	37	40	44	
B*40012	2	7	9	14	17	18	19	22	30	32	33	37	40	44	
B*4002	2		9	14	17	18	19	22	30	32	33	37	40	44	45
B*4003	2		9	14	17	18	19	30	32	33	34	37	40	44	45
B*4004	2	6	10	11	14	17	18	19	22	30	32	33	37	44	
B*4005	2	5	9	14	17	18	19	22	30	33	37	40	44	45	
B*4006	2		9	14	17	18	19	22	26	30	32	33	37	44	45
B*4007	2	7	9	14	17	18	19	22	30	32	37	40	44		
B*4101	2	6	9	17	19	22	25	26	30	32	33	37	44		
B*4102	2	9	17	19	22	25	30	33	37	40	44				
B*4201	2	9	17	19	22	24	25	30	31	32	35	37	40	44	
B*4402	2	5	9	10	19	32	33	37	43	45					
B*44031	2	5	9	10	18	19	32	33	37	43	45				
B*44032	2	5	6	10	18	19	26	32	33	43	45				
B*4404	2	9	10	19	25	32	33	37	42	43	45				
B*4405	2	5	9	10	32	33	37	45							
B*4406	2	4	5	9	19	24	32	37	43	45					
B*4501	2	5	6	11	19	20	30	32	33	37	43	45			
B*4601	2	5	7	9	14	15	17	19	30	34	37	40	45		

Allele															
B*4701	2	7	14	9	14	17	18	19	32	33	37	40	43	45	
B*4801	2	9	17	17	18	19	22	30	31	32	33	37	40	44	
B*4802	2	5	6	10	11	17	18	19	20	30	31	32	33	34	37
B*4901	2	4	5	6	11	17	18	19	20	20	33	37	45		
B*5001	2	5	6	11	17	18	19	20	30	33	33	37	45		
B*5101	2	4	5	9	17	18	19	20	22	24	26	33	37		
B*5102	2	4	5	9	17	18	19	22	24	26	26	37	45		
B*5103	2	4	5	9	18	19	22	24	26	26	37				
B*5104	2	4	5	9	18	17	18	22	24	26	45				
B*5105	2	4	5	9	17	19	22	24	26	32	37	42	45		
B*52011	2	4	5	9	17	18	19	22	26	33	45				
B*52012	2	4	5	9	17	18	19	22	26	33	45				
B*5301	2	4	5	6	10	11	17	18	19	20	24	32	34	37	
B*5401	2	11	17	18	19	22	24	25	30	32	35	36	37	45	
B*5501	2	11	17	18	19	22	24	25	30	35	35	37	45		
B*5502	2	11	17	18	19	22	24	25	30	32	35	37	45		
B*5601	2	5	11	17	18	22	24	24	30	32	35	37	45		
B*5602	2	5	7	17	14	17	18	19	24	26	32	35	37	40	45
B*5701	2	4	5	9	14	17	18	22	26	32	32	34	37	45	
B*5702	2	4	5	9	14	17	18	22	26	32	32	37	42	45	
B*5703	2	4	5	9	10	17	18	19	22	26	32	37	45		
B*5801	2	4	5	6	17	18	19	20	20	32	34	37	45		
B*5802	2	4	5	11	17	18	19	20	32	34	34	37	40	45	
B*5803	2	4	5	11	17	18	19	20	32	34	34	37	45		
B*5901	2	4	11	17	18	22	24	25	32	37	37	45			

B*67011	2	7	9	17	22	23	24	25	30	31	32	35	37	40	45
B*67012	2	7	9	17	19	22	23	24	25	30	31	32	35	37	40
B*7301	14	17	18	19	23	24	30	32	44	45					
B*7801	2	5	9	17	18	19	22	24	26	30	41	45			
B*7802	2	5	9	17	18	19	22	24	26	30	45				
B*8101	2	9	14	17	18	19	22	24	30	31	32	35	37	40	44

HLA-C Alleles

Cw*0101	17	19	20	25	30	36	37	40	42	44			
Cw*0102	17	19	20	25	30	36	37	40	42	44	45		
Cw*0201	7	15	17	19	20	34	37	40	44				
Cw*02021	7	15	17	19	20	34	36	37	40	44	45		
Cw*02022	7	15	17	19	20	34	36	37	40	44	45		
Cw*0302	2	5	6	14	17	18	19	20	26	30	34	36	45
Cw*0303	2	5	6	10	14	17	18	19	20	26	30	36	45
Cw*0304	2	5	6	10	14	17	18	19	20	26	30	36	45
Cw*0401	2	7	8	14	17	19	23	25	37	40	42	45	
Cw*0402	2	7	8	14	17	19	23	25	37	40	45		
Cw*0501	6	7	11	19	20	21	23	25	36	40	42	44	
Cw*0602	11	14	15	17	19	20	25	34	37	40	44	45	
Cw*0701	6	7	11	17	18	19	20	25	30	34	36	37	40
Cw*0702	7	17	18	19	20	30	34	36	40	44	45		
Cw*0703	7	17	18	19	20	30	34	36	37	40	44	45	

	2	6	7	10	11	14	15	17	18	19	20	23	25	30	34	36	37	40	42	44	45
Cw*0704		6			11			17		19	20	23	25	30			37			44	
Cw*0801		6	7		11	14		17	18	19	20	23	25	30		36	37	40			
Cw*0802		6	7		11			17		19	20	23	25	30		36	37	40	42	44	
Cw*0803		6	7		11	14		17	18	19	20	23	25	30		36		40		44	
Cw*1201			7			14	15	17		19	23		25	30			37	40			45
Cw*12021			7			14	15	17		19	20		25	30	34	36		40		44	45
Cw*12022			7			14	15	17		19	20		25	30	34	36	37	40		44	45
Cw*1203					11	14	15	17		19	20		25	30	34	36	37	40		44	45
Cw*1301			7			14	15	17		19	20		25	30	34		37	40		44	45
Cw*1401	2					14		17		19	20		25	30	34	36	37	40	42		
Cw*1402	2					14		17		19	20		25	30	34	36	37	40	42		45
Cw*1403	2					14		17		19	20		25	30	34	36	37	40	42		45
Cw*1501		6		10	11	14		17	18	19	20		25			36	37				45
Cw*1502		6		10	11	14		17	18	19	20		25			36	37			44	45
Cw*1503		6		10	11	14		17	18	19	20		25			36	37			44	45
Cw*1504		6		10	11	14		17	18	19	20		25		34	36	37			44	45
Cw*1505		6		10	11	14		17	18	19	20	23	25			36	37			44	45
Cw*1601					11			17		19	20		25	30	34	36		40			45
Cw*1602					11			17		19	20		25		34	36	37	40			45
Cw*1603					11	14		17		19	20		25	30	34	36	37			44	45
Cw*1701		6			11	14		17	18	19	20	23				36					45

Table 3

The location of the target sequences for URSTO probes in HLA class I genes

Sequence range	No. of probes	Probe identity
1. 92-109	1	31
2. 113-130	1	2
3. 212-229	1	36
4. 236-299	7	24,28,19,35,41,3,33 (a)
5. 293-310	1	29
6. 307-324	2	4,30
7. 333-430	14	6,8,10,26,11,20,7,40,9 (b)
		23,12,34,43,22
8. 454-471	1	44
9. 478-495	1	14
10. 520-578	9	5,25,32,15,16,13,17,18,42 (c)
11. 580-597	1	37
12. 597-614	1	45

The probe identity numbers are same as in Table 1; a, b and c indicate the hypervariable sequence regions which are present in the three loci. The sequence range refers to the base positions in exons 2 and 3.

The panel of probes used in the invention generally consists of a sufficient number of probes to uniquely identify the majority of the alleles of the polyallelic gene. For example, the panel of probes may be selected so

5 as to uniquely identify at least 50%, at least 75%, at least 90%, at least 95%, at least 99% or 100% of the alleles in the polyallelic system. The exact number of probes will vary depending on the gene, but is typically from 10 to 100, preferably from 20 to 70 or from 30 to 50.

10 Each probe may recognise a sequence motif that is present in, for example, from 2 to 30, from 2 to 20, for 4 to 20 or from 6 to 16 alleles of the polyallelic gene.

When the invention is applied to the HLA class I system,
15 the panel of probes preferably comprises from 20 to 40 probes which each recognises the motif recognised by one of the probes set out in Table 1. Each of the probes in the panel may have a sequence at least 40%, at least 60%, at least 80%, at least 90% or at least 95% identical to either

20 (i) a sequence of one of the probes set out in Table 1 or (ii) a sequence complementary to a sequence of one of the probes set out in Table 1. A probe in the panel may have a sequence that is shifted along the HLA class I gene sequence by a certain number of nucleotides compared to a

25 probe set out in Table 1; for example, a probe may be shifted along by from 1 to 10 nucleotides (e.g. from 1 to 5 nucleotides) in either a 5' or a 3' direction compared to a probe set out in Table 1.

30 The probes used in the invention may be labelled with any one of a variety of detectable labels in order to facilitate their detection. Examples of suitable labels include digoxygenin (which may be detected using an anti-digoxygenin antibody coupled to alkaline phosphatase),
35 radiolabels, biotin (which may be detected by avidin or streptavidin conjugated to peroxidase) and fluorescent labels (e.g. fluorescein and rhodamine).

- 32 -

The kit according to the invention may comprise a database which indicates which probes in the panel in the kit recognise each allele of the polyallelic gene. The database may be a paper database or a computer database.

5   The database may be compiled by examining the sequences of the allele of the polyallelic genes and noting the probes which would be expected to bind to each allele. The accuracy of a database compiled by such a technique may be verified by experimentally determining which alleles are

10 bound by each probe in the panel. Table 2 contains a database showing which of the 40 probes in Table 1 bind specific HLA class I alleles. The kit may also contain one or more known alleles as control(s). Such controls can be used to verify that an experiment carried out using the kit

15 has worked correctly.

It is highly desirable that the sample of allele used in the invention comprises one allele only and is not contaminated by other alleles of the same gene. The

20 presence of two similar alleles in the sample can give confusing results and prevent conclusive identification of the alleles. Individuals are often heterozygous with respect to the alleles of a particular gene, i.e. individuals often have two different alleles of the same

25 gene, and these alleles normally need to be separated before carrying out the invention.

In view of the fact that the difference between two alleles of a gene can be as little as one nucleotide, it is often

30 difficult to separate the alleles from a mixture of the alleles. These difficulties are increased in genes which have a very large number of different alleles, such as the major histocompatibility complex (MHC) genes (e.g. the human leucocyte antigen (HLA) class I genes which have 222

35 known alleles).

A new method for separating alleles of a gene from a

- 33 -

mixture of alleles has now been found, which is referred to
herein as "Complementary Strands Analysis" (CSA). This
method is described in detail in an international (PCT)
patent application in the name of the Anthony Nolan Bone
5   Marrow Trust being filed on the same day as this
application. The method comprises

    (i)   amplifying the alleles in the mixture of alleles;
    (ii) hybridising single strands of the amplified
10   alleles with a complementary strand of a
           reference allele to form duplexes; and
    (iii) separating the duplexes.

The different alleles in the original mixture give rise to
15   duplexes having different numbers of mismatches compared to
a selected complementary reference DNA strand. This allows
the duplexes to be separated by, for example, gel
electrophoresis. The separated duplexes can then be
analysed by the method of the invention to identify the
20   alleles that were present in the original mixture.

A preferred form of the CSA method comprises

    (i)   amplifying the mixture of alleles employing a
25   pair of primers in which one of the primers
           carries a ligand, so as to produce an amplified
           mixture of double-stranded alleles in which one
           of the strands carries a ligand;

30   (ii) contacting the amplified mixture of double-
           stranded alleles with a receptor on a solid
           support under conditions such that the ligand
           binds to the receptor;

35   (iii) separating the mixture of double-stranded alleles
           into single-strands and removing the strands that
           are not bound to the support by the ligand;

(iv) recovering the remaining strands from the support;

(v) mixing the recovered strands with a complementary strand of a reference allele so as to form duplexes; and

(vi) separating the duplexes.

Another form of the CSA method comprises

(i) amplifying the alleles in the mixture employing a pair of primers in which one of the primers carries a high molecular weight molecule, so as to produce an amplified mixture of double-stranded alleles in which one of the strands carries a high molecular weight molecule;

(ii) separating the mixture of double-stranded alleles into single strands;

(iii) mixing the single strands with a complementary strand of a reference allele so as to form duplexes; and

(iv) separating the duplexes.

This form of CSA overcomes the need for solid support systems by conjugating one primer of a pair of primers directly to a high molecular weight molecule (e.g. a protein). The amplified product after hybridisation can be applied directly to a separating gel. The high molecular weight conjugates are retained in the gel compared to the duplex without attachment of the high molecular weight molecule.

In yet another form of the CSA method, there is provided a

- 35 -

method for separating an allele from a mixture of alleles, which method comprises

(i)   amplifying a single strand of each of the alleles in the mixture;

(ii) mixing the amplified single strands with a complementary strand of a reference allele so as to form duplexes; and

(iii) separating the duplexes.

The amplification of the single strand can be done, for example, by asymmetric PCR.

This form of CSA overcomes the need for both solid support systems and conjugation of one primer of a pair to a high molecular weight molecule. However, in the embodiment it is possible to use a primer carrying a ligand such as a hapten in order to facilitate capture of the amplified strand with a receptor such as an antibody and separation of the amplified strand from other components in the amplification mixture.

In each of the above forms of CSA, the reference allele may be provided in single-stranded form by essentially the same steps as used to provide the test alleles in single-stranded form.

The CSA method provides an improvement over prior methods for separating alleles. The advantages offered by CSA can be summarised as follows:

(a)   The method provides a high resolution between different alleles and differences of as little as one nucleotide between alleles can be detected.

- 36 -

(b)   The method allows simultaneous and rapid
        processing of a large number of samples.

(c)   The method is comparatively inexpensive to
5   perform, particularly when compared to prior
        methods which achieve a high level of resolution.

(d)   The method uses techniques that can be performed
        easily without recourse to complex and expensive
10   technology.

The reference allele used in the CSA method generally has a
known sequence. Further, the reference allele is usually
chosen so as to have a similar allotype to an allotype that
15   at least one of the test alleles is suspected of having.
For example, it may be known that a test allele is of the
HLA-A02 type from serological data, but it may not be known
which of the seventeen A02 sub-types the allele is.   In
this case, the reference allele may be chosen to be of sub-
20   type A0201 and the method of the present invention could
then be used to determine which of the A02 sub-types the
test allele is.

The reference strand may be obtained from (a) a homozygous
25   source, (b) a heterozygous source from which individual
strands are isolated by gel separation after amplification
steps or (c) DNA synthesis. There are now about 500
internationally recognised cell lines which contain HLA
alleles of known sub-type and these cell lines can be used
30   as a source of reference alleles.

In the CSA method, the amplification steps may be carried
out by polymerase chain reaction (PCR).

35   The ligand/receptor system used in the CSA method may, for
example, be the biotin/streptavidin system. Direct
conjugation of the primer via a linking group, such as

short poly A, to the beads is an alternative. When the
biotin/streptavidin system is used, one of the primers used
in each of the amplification steps may be labelled with
biotin, so that when the amplification reaction is carried
5   out double-stranded DNA is produced in which one strand
carries a biotin label. The double-stranded DNA may then
be bound to a solid support coated with streptavidin.

The solid support used in the CSA method is typically
10 magnetic beads. However, other supports may be used, such
as the matrix of an affinity chromatography column. When
the support is in the form of magnetic beads, the two
strands of the amplified DNA are separated by attracting
the beads to a magnet and washing the beads under
15 conditions such that the double-stranded DNA dissociates
into single-strands. The dissociation is typically
performed by incubating the beads (e.g. three times) under
alkaline conditions (e.g. 0.1 M or 0.15 M NaOH) at room
temperature for about 5 or 10 minutes. Usually, the strand
20 which is not bound to the support by the ligand is then
discarded, although it is equally possible to retain the
strand that is not bound to the support and discard the
strand that is bound to the support.

25 The strand that remains attached to the support may be
recovered from the support by incubating the support under
conditions such that the ligand/receptor complex
dissociates. When the biotin/streptavidin system is used,
the support is typically heated to e.g. 95°C for about 5
30 minutes; this ensures denaturation of the streptavidin
molecule to release the biotinylated single strand which is
then recovered.

At this stage, there have been provided a single-stranded
35 unknown allele and the complementary strand of a reference
allele. The two strands are then mixed together under
conditions in which they hybridise to form duplexes.

- 38 -

Typically, the hybridisation step is performed by heating the mixture of strands at about 95°C for about 3 min, at about 70°C for about 5 min and then at about 65°C for about 45 min.

5

Under these conditions, duplexes are formed which can subsequently be separated by gel electrophoresis (e.g. polyacrylamide gel electrophoresis). The electrophoresis may be carried out under denaturing or non-denaturing
10   conditions. The use of denaturing conditions may enhance separation.

As an alternative separation technique to gel electrophoresis, high pressure liquid chromatography (HPLC)
15   may be used.

In the embodiment of the CSA method in which one of the pair or primers is conjugated to a high molecular weight molecule, the molecule may be a protein such as bovine
20   serum albumin (BSA). The molecular weight of the high molecular weight molecule is such that it causes the DNA molecule to which it is attached to be sufficiently retarded in the separation step (e.g. the electrophoresis step) to allow the DNA molecule to be separated from a
25   duplex without a high molecular weight compound attached. For example, the molecular weight of the high molecular weight molecule may be from 10 to 200 kDa, preferably 20 to 100 kDa.

30   The invention may be used to match a prospective donor in a tissue or organ transplant operation with a prospective recipient. In particular, the invention may be used to identify the alleles of the prospective recipient and donor, and hence to determine whether they have compatible
35   alleles. The prospective recipient and donor may, for example, be a prospective recipient and donor in a bone marrow or kidney transplant operation.

Other proposed uses of the invention include determination of the paternity of an individual by identifying one (or more) of his alleles to see if it is the same as a corresponding allele of a putative father. The invention
5 may also be used in forensic medicine to determine the origin of a sample of body tissue or fluid, as a follow up technique in treatment of haematological malignancies or inherited disorders, in adoptive immunotherapy, and in identification of bacteria and viruses.
10

The following example illustrates the invention.

EXAMPLE

15 **METHODS**

1.- Locus specific amplification of HLA class I genes

For typing purposes, amplification of exons 2 and 3 is
20 desirable, and the primers were therefore selected to amplify the stretch of the genome between intron 1 and intron 3. The localisation and nucleotide sequences of the HLA locus-specific primers used are given in the reagents section.
25

PCR reactions were performed in a total volume of 100 μl using 1μg of genomic DNA and 25 pmoles of each locus-specific primer. The 3' primer was biotinylated at the 5' end. This arrangement ensures the incorporation of the
30 biotinylated primer onto the amplified antisense DNA strand. PCR conditions are given in the following table.

- 40 -

Thermocycling conditions
A, B and Cw loci

95° C	4 min.	1 cycle

95° C	30 sec.	
65° C	50 sec.	33 cycles
72° C	30 sec.	
72° C	8 min.	1 cycle

2.- Separation of the amplified DNA strands
a) Removal of non-biotinylated strand:

Magnetic beads with covalently coupled streptavidin on the
surface were added to the PCR product and incubated for 30
minutes at 43° C. In this way the amplified PCR product was
immobilised by the interaction of biotin and streptavidin.
After incubation, the tubes were placed against a magnet
and the beads were washed with washing buffer to remove the
remaining PCR reaction components.

The non-biotinylated DNA strand was then dissociated from
the beads by incubation with 0.1 M NaOH at room temperature
(r.t.) for 10 minutes. Following this the beads were
washed to remove excess NaOH and resuspended in 50 μl of
hybridisation buffer.

b) Removal of biotinylated DNA strand:

The bead suspension was heated at 95°C for 5 minutes, which
ensures denaturation of the streptavidin molecule to
release the biotinylated amplified anti-sense single strand
which was then removed and placed in a clean tube.

At this stage, the isolates contained single biotinylated
DNA strands from each allele.

3.- Hybridisation with locus specific reference sense single strand DNA

The biotinylated anti-sense strand(s) from above were mixed
5   with a locus specific reference sense strand, Rf A, Rf B
      and Rf C for HLA-A, B and C respectively (see below), and
      the mixture was heated at 95°C for 3 min., incubated at
      70°C for 5 min. and then at 65°C for 45 min. Under these
      conditions the sense and anti-sense strands were
10   hybridised. The heteroduplexes formed by each allele
      antisense strand with the locus specific reference sense
      strand could subsequently be separated from each other by
      electrophoresis in non-denaturing polyacrylamide gel.

15

4.- Preparation of locus specific reference sense single strand DNA.

DNA was extracted from three homozygous 10th IHW cell
20   lines. The following cell lines were selected as locus
      specific reference DNA: STEINLIN (HLA-A*0101), SP0010 (HLA-
      B*4402) and STIENLIN (HLA-Cw*0701).

The PCR conditions for amplification were as above, with
25   the exception that in each case the locus-specific 5'
      primer was biotinylated (5' end). The PCR products were
      analysed by PAGE to assess the fidelity of the
      amplification and in all cases a single band was obtained.

30   The biotinylated single sense strand was isolated as
      described above and its purity was tested by a
      heating/annealing cycle of the sample followed by agarose
      electrophoresis. In each case only a single band of the
      expected size was observed.

35

5.- Separation of alleles

This step is critical for the separation of the allelic products from heterozygous subjects and from coamplified
5   non-specific products.

The heteroduplexes formed as described in step 3 were separated from each other by electrophoresis analysis which was performed on an 8% non-denaturing polyacrylamide gel at
10   room temperature (200 volts for 6 hours). The DNA was visualised by ethidium bromide staining and U.V. light.

The heteroduplexes from heterozygote individuals were resolved into two bands, while DNA from homozygote subjects
15   produced a single band.

6.- Identification of HLA class I alleles

The bands were excised from the gel from which the DNA was
20   eluted and blotted on the same membrane for three loci.
For heterozygous subjects two dots per locus were prepared.
Several subject samples were blotted on each membrane.
According to the number of the URSTO probes several membranes were prepared.
25

The oligonucleotide probes were labelled with digoxigenin (DIG) at the 3' end (Boehringer, according to specification).

30   Hybridisation and washing solutions contained TMACl (3M), and membranes were hybridised at 54°C and washed at 58°C.
Oligonucleotide binding was detected by chemiluminescence; anti-DIG-antibody conjugated to alkaline phosphatase and CSPD were added to membranes. After incubation the
35   chemiluminescence was detected by X-ray films.

RESULTS

A.- Separation of alleles by CSA

5   Anti-sense strands from more than 20 samples were isolated,
hybridised with the HLA-A locus specific reference sense
strand (STEINLIN A*0101), and analysed by 8% non-denaturing
polyacrylamide gel electrophoresis. In all cases there was
a good correlation between the number of bands observed and
10   the zygosity of the sample. For example, two bands were
seen for heterozygous samples whereas one band was seen for
homozygous samples. These bands were always observed in
the same area of the gel, the lower half nearest to the
anode. A representative autoradiograph from CSA analysis
15   is presented in Figure 2B.

B.- Identification of alleles by URSTO

For initial testing of the labelled URSTO probes, DNA from
20   10 homozygous EBV transformed B cell lines set out in the
following table was amplified for the HLA-A, B and C loci
using locus specific primers and blotted onto 10 nylon
membranes.

25

- 44 -

Specificity of the 10 homozygous cell lines used in URSTO

STEINLIN	A*0101	B8	C*0701
LBUF	A*3001	B*1302	C*0601
BM14	A3	B7	C7
JBUSH	A32	B38	C*1203
BTB	A2	B27	C1
WT47	A32	B44	C5
SWEIGOO7	A*2902	B*4002	C*02022
BM92	A*2501	B*5101	C1
SPL	A31	B62	C1
SPOO10	A2	B*4402	C5

Four URSTO probes (P3, P4, P5 and P29) labelled with DIG
were then hybridised to these membranes under specific
conditions: 54°C for 90 minutes in TMACl solution. The
membranes were then washed (X3) under stringent conditions:
58°C for 10 min. in TMACl wash solution. Detection was
performed by anti-DIG alkaline-phosphatase conjugate,
followed by development with CSPD.

It was found that the four URSTO probes gave the expected
pattern (see the following table).

Results of 4 URSTO probes

Allelic specificity	URSTO probes			
	P3	P4	P5	P29
A*0101	−	−	−	−
B8	−	−	−	−
C*0701	−	−	−	−
A*3001	−	−	−	+
B*1302	−	−	−	−
C*0601	−	−	−	−
A3	−	−	−	+
B7	−	−	−	−
C7	−	−	−	−
A32	−	+	−	−
B38	−	+	−	−
C*1203	−	−	−	−
A2	−	−	−	+
B27	−	−	−	−
C1	−	−	−	−
A32	−	+	−	−
B44	−	−	+	−
C5	−	−	−	−
A*2902	−	−	−	−
B*4002	−	−	−	−
C*02022	−	−	−	−
A*2501	−	+	−	−
B5101	−	+	+	−
C1	−	−	−	−
A31	+	−	−	+
B62	−	−	+	−
C1	−	−	−	−
A2	−	−	−	+
B*4402	−	−	+	−
C5	−	−	−	−

Tests with all 40 probes and a large number of
internationally defined samples (International
Histocompatibility workshop cell lines) indicate that each
allele tested gave the pattern shown in Table 2 (see

5   Figures 3 and 4).

A list of HLA class I alleles which have been isolated and
identified by URSTO is set out in the following table:

HLA class I alleles which have been isolated
5   and identified by URSTO method

HLA-A (n=33)

A*0101, A*0102, A*0201, A*0202, A*0203, A*0204, A*0205,
10 A*0206, A*0207, A*0208, A*0209, A*0210, A*0211, A*0212,
A*0213, A*0216, A*0217, A*0301, A*1101, A*2301, A*2402,
A*2403, A*2501, A*2601, A*2902, A*3001, A*3002, A*3101,
A*3201, A*3301, A*6601, A*6602, A*6802

15 **HLA-B (n=30)**

B*0702, B*0801, B*1302, B*1402, B*1501, B*1502, B*1520,
B*1801, B*3501, B*3701, B*3801, B*4001, B*4002, B*4101,
B*4201, B*4402, B*4403, B*4601, B*4701, B*4801, B*4901,
B*5001, B*5101, B*5201, B*5301, B*5502, B*5701, B*5801,
20 B*5802, B*6701

HLA-Cw (n=18)

Cw*0102, Cw*0202, Cw*0302, Cw*0303, Cw*0304, Cw*0401,
Cw*0501, Cw*0602, Cw*0701, Cw*0702, Cw*0704, Cw*0802,
25 Cw*1202, Cw*1203, Cw*1402, Cw*1502, Cw*1601, Cw*1701

Number of different heterozygous combinations tested: HLA-A
19, HLA-B 14, and HLA-Cw 11.
30

In the identification of the alleles set out in the above
table, DNA was extracted from 63 B-lymphoblastoid cell
lines; these included 22 heterozygous and 41 homozygous
lines. After PCR amplification with locus specific primers
35 as described above the anti-sense single strands were
isolated and hybridised as described above with the
appropriate reference strands (A*0101, B*4402, Cw*0701).

The allelic bands were resolved in non-denaturing PAGE and
eluted from low melting point agarose as described above.
The DNA from the isolated bands was blotted on 40 nylon
membranes (as in Figure 3) and hybridised with URSTO
5   probes.   Alleles were identified by comparison of patterns
with those in Table 2.

REAGENTS

10 **A) Nucleotide sequences of primers used for locus-specific
    amplification:**

     5' A locus primer: GAA ACG/C GCC TCT GT/CG GGG AGA
                                 AGC AA

15   (Intron 1: 21-46)

     3' A locus primer: TGT TGG TCC CAA TTG TCT CCC CTC
         (Intron 3: 66-89)

20 5' B locus primer: GGG AGG AGC GAG GGG ACC G/CCA G
         (Intron 1: 36-57)

     3' B locus primer: GGA GGC CAT CCC CGG CGA CCT AT
         (Intron 3: 37-59)

25
     5' C locus primer: AGC GAG GG/TG CCC GCC CGG CGA
         (Intron 1: 42- 61)

     3' C locus primer: GGA GAT GGG GAA GGC TCC CCA CT
30   (Intron 3: 12-35)

     B) Buffers:

35 Washing buffer: 10 mM   Tris-HCl pH 7.5
                                 1.0 mM   EDTA
                                 2.0 M   NaCl

Hybridisation buffer:	20 mM	Tris-HCl pH 8.4
	50 mM	KCl

PCR buffer:	20 mM	Tris-HCl pH 8.4
	50 mM	KCl
	0.2 mM	MgCl2

TE buffer	10 mM	Tris-HCl pH 7.5
	1 mM	EDTA

C) Various

Dynabeads M-280 Streptavidin (10 mg/ml)

Magnetic particle concentrator -Dynal MPC

Nylon membranes, positively charged (Boehringer Mannheim)

CSPD-Disodium3-(4-methoxyspiro{1,2-dioxetane-3,2'-(5'cholo)t ricyclo[3.3.1.13,7] decan}-4-yl)   Boehringer Mannheim

DIG Oligonucleotide 3' -End Labeling Kit (Boehringer Mannheim)

Anti-digoxigenin-AP Fab fragments (Boehringer Mannheim)

A Thermal cycler (PTC-200 Peltier Thermal Cycler MJ Research)

Ultrapure dNTP set, 2' -Deoxynucleoside 5' -Triphosphate (Pharmacia Biotech)

Taq DNA Polymerase (Gibco BRL)

50 mM MgCl2

- 50 -

0.1 M NaOH

SeaPlaque Agarose (Flowgen Instruments Ltd)

5   Protogel, 30% Acrylamide and 0.8% Bisacrylamide (National
Diagnostics)

REFERENCES

10 1.- D. Weisdorf, R. Haake & B. Blazar. Risk factors for
acute graft-versus-host disease in histocompatible donor
bone marrow transplantation. *Transplantation* 1991: 51:
1197-1203

15 2.- L. A. Ochs, W. J. Miller, A. H. Filipovich, R. J.
Haake, P. B. McGlave, B. R. Blazar, N. K. C. Ramsay, J. H.
Kersey & D. J. Weisdorf. Predictive factors for chronic
graft-versus-host disease after histocompatible sibling
donor bone marrow transplantation. *Bone Marrow Transplant*
20 1994: 13; 455-460

3.- L. A Smyth, C. S. Witt, F. T. Christiansen, R. P.
Hermann, P. N. Hollingsworth, D. C. Townend, E Edward & R.
L. Dawking. The MHC influences acute graft versus host
25 disease in MHC matched adults undergoing allogeneic one
marrow transplantation. *Bone Marrow Transplant* 1993: 12;
351-355

4.- A. Bacigalupo, P. Gualandi, M. T. Van Lint, M.
30 Sessarego, F. Frassoni, D. Occhini, T. Lamparelli, R.
Oneto, V. Vitale, R. Corvo, E. Raul de la Torre & A. M.
Marmont. Multivariate analysis of risk factors for survival
and relapse in chronic granulocytic leukemia following
allogeneic marrow transplantation: impact of disease
35 related variables (Sokal score). *Bone Marrow Transplant*
1993: 12; 443-448

5.- J. Sierra, A. Graneda, J. Garcia, A. Valls, E. Carreras, M. Rovira, C. Canals, E. Martinez, C. Punti, M. Algara, P. Martin, A. Merino, M. J. Terol, A. Urbano-Ispizua & C. Rozman. Autologous bone marrow transplantation for acute leukemia: results and prognostic factors in 90 consecutive patients. *Bone Marrow Transplant* 1993: 12; 517-523

6.- J. L. Bidwell. Applications of the polymerase chain reaction to HLA class II typing. *Vox Sang* 1992: 63; 81-89

7.- P. Parham, E. Adams & k. L. Arnett. The Origins of HLA-A, B, C Polymorphism. *Immunological Reviews* 1995: 143; 141-180

8.- P. Krausa, M. Brywka III, D. Savage, K. M. Hui, M. Bunce, J. L. F. Ngai, D. L. T. Teo, Y. W. Ong, D. Barouch, C. E. M. Allsop, A. V. S. Hill, A. J. McMichael, J. G. Bodmer & M. J. Browning. Genetic polymorphism within HLA-A*02: significant allelic variation revealed in different populations. *Tissue Antigens* 1995: 45; 223-231

9.- P. Krausa, J. G. Bodmer, M. J. Browning. Defining the common subtypes of HLA A9, A10, A28 and A19 by use of ARMS/PCR. *Tissue Antigens* 1993: 42; 91-99

10.- C. W. Summers, V. J. Hampson & G. M. Taylor. HLA class I non-coding nucleotide sequences, 1992. *European Journal of Immunogenetics* 1993: 20; 201-240

11.- S. Hoshino, A. Kimura, Y. Fukuda, K. Dohi & T. Sasazuki. Polymerase chain reaction-single-strand conformation polymorphism analysis of polymorphism in DPA1 and DPB1 genes: A simple, economical, and rapid method for histocompatibility testing. *Human Immunology* 1992: 33; 98-107

- 52 -

12.- M. Orita, Y. Suzuki, T. Sekiya & K. Hayashi. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 1989: 5; 874-879

5

13.- Y. Suzuki, T. Sekiya & K. Hayashi. Allele-specific polymerase chain reaction: A method for amplification and sequence determination of a single component among a mixture of sequence variants. *Analytical Biochemistry* 1991: 10 192; 82-84

14.- R. Blasczyk, U. Hahn, J. Wehling, D. Huhn & A. Salama. Complete subtyping of the HLA-A locus by sequence-specific amplification followed by direct sequencing or single- 15 strand conformation polymorphism analysis. *Tissue Antigens* 1995: 46; 86-95

15.- R. Blasczyk, J. Wehling, B. S. Kubens, U. Hahn, D. Huhn & A. Salama. A novel HLA-A24 allele (A*2405) 20 identified by single-strand conformation polymorphism analysis and confirmed by solid-phase sequencing and isoelectric focusing. *Tissue Antigens* 1995: 46; 54-58

16.- M.Orita, H.Iwahana, H. Kanazawa, K.Hayashi and 25 T.Sekiya. Detection of polymorphism of human DNA by gel electrophoresis as single-strand conformation polymorphism. *Proc Natl Acad Sci USA* 1989: 86; 2766-2770

17.- G. Fischer and L. S Lerman. DNA fragments differing 30 by a single base-pair substitution are separated in denaturing gradient gels: Correspondence with melting theory. *Proc Natl Acad Sci USA* 1983: 80; 1579-1583

18.- K. Henco and M. Heibey. Quantitative PCR: the 35 determination of template number copy numbers by temperature gradient gel electrophoresis. *Nucleic Acid Res* 1990: 18; 6733-6734

19.- T. M. Clay, J. L. Bidwell, M. R. Howard & B. A. Bradley. PCR-fingerprinting for selection of HLA matched unrelated marrow donors. *Lancet* 1991: 337; 1049-52

20.- A. Bhattacharyya & D. M. J. Lilley. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res* 1989: 17; 6821-6840

21.- F. Aboul-ela, D. Koh & I. Tinoco. Base-base mismatches. Thermodynamics of double helix formation for dCA3XA3G + dCT3YT3G (X, Y = A, C, G, T) *Nucleic Acids Res* 1985: 13; 4811-4824

22.- J. Y. Tong, A. Hammad, W. A. Rudert, M. Trucco & S. Hsia. Heteroduplexes for HLA DQB1 identity of family members and kidney donor-recipient pairs. *Transplantation* 1994: 57; 741-745

23.- R. Sorrentino, I. Cascino & R. Tosi. Subgrouping of DR4 alleles by DNA heteroduplex analysis. *Human Immunology* 1992: 33; 18-23

24.- M. White, M. Carvalho, D. Derse, S. O'Brien & M. Dean. Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 1992: 12; 301-306

25.- M. Carrington, M. White, M. Dean, D. Mann & F. E. Ward. The use of DNA heteroduplex patterns to map recombination within the HLA class II region. *Human Immunology* 1992: 33; 114-121

26.- N. A. P. Wood, T. M. Clay & J. L. Bidwell. HLA-DR/Dw matching by PCR Fingerprinting: The origin of PCR fingerprints and further applications. *European Journal of Immunogenetics* 1991: 18; 147-153

- 54 -

27.- M. Allen, L. Liu & U. Gyllensten. A comprehensive polymerase chain reaction-oligonucleotide typing system for HLA class I A locus. *Human Immunology* 1994: 40; 25-32

28.- X. Gao, I. B. Jakobsen & S. W. Serjeantson. Characterization of the HLA-A polymorphism by locus-specific polymerase chain reaction amplification and oligonucleotide hybridization. *Human Immunology* 1994: 41; 267-279

29.- A. Selvakumar, C. B. Granja, M. Salazar, S. M. Alosco, E. J. Yunis & B. Dupont. A novel subtype of A2 (A*0217) isolated from the South American Indian B-cell line AMALA. *Tissue Antigens* 1995: 45; 343-347

30.- E. W. Petersdorf & J. A. Hansen. A comprehensive approach for typing the alleles of the HLA-B locus by automated sequencing. *Tissue Antigens* 1995: 46; 73-85

31.- J. A. Madrigal, M. P. Belich, W. H. Hildebrand, R. J. Benjamin, A. M. Little, J. Zemmour, P. D. Ennis, F. E. Ward, M. L. Petzl-Erler, E. D. du Toit & P. Parham.Distinctive HLA-A, B antigens of black populations formed by interallelic conversion. *J. Immunol* 1992: 149; 3411-3415

32.- J. Zemmour & P. Parham. HLA Class I Nucleotide Sequences, 1992. *European Journal of Immunogenetics* 1993: 20: 29-45.

33.- N. Cereb, P. Maye, S. Lee, Y. Kong & S. Y. Yang. Locus-specific amplification of HLA class I genes from genomic DNA. *Tissue Antigens* 1995:

34.- J. A. Madrigal, W. H. Hildebrand, M. P. Belich, R. J. Benjamin, A. M. Little, J. Zemmour, P. D. Ennis, F. E. Ward, M. L. Petzl-Erler, E. D. du Toit & P. Parham.

Structural diversity in the HLA-A10 family of alleles:
Correlations with serology. *Tissue Antigens* 1993: 41; 72-80

35.- D. Chen, W. Endres, S. A. Meyer & W. Stangel. A
polymerase chain reaction-sequence-specific oligonucleotide
procedure for HLA class II typig using biotin- and
digoxigenin-labelled probes simultaneously in
hybridization. *Human Immunology* 1994: 39; 25-30

36.- GE Hawes, L Struyk, PJ van den Elsen. Differential
usage of T cell receptor V gene segments in CD4⁺ and CD8⁺
subsets of T lymphocytes in monozygotic twins, J Immunol
1993:150; 2033-2045

37.- C Giachino, MP Rocci, G De Libero, G Oderda, N Ansaldi
and N Migone. An alternative approach to the assessment of
T cell clonality in celiac disease intestinal lesions
through cDNA heteroduplex analysis of T-cell receptor VJ
junctions. Human Immunology 1994: 40; 303-311

38.- R. Arguello, H. Avakian, J.H. Goldman and J.A.
Madrigal. A novel method for simultaneous high resolution
identification of HLA-A, HLA-B and HLA-Cw alleles. *Proc
Natl Acad Sci. USA* 1996: 93; 10961-10965

CLAIMS

1.   A method for identifying an unknown allele of a
polyallelic gene, which method comprises

5   (i)   contacting the unknown allele with a panel of
                 probes, each of which recognises a sequence motif
                 that is present in some alleles of the
                 polyallelic gene but not in others;

10   (ii) observing which probes recognise the unknown
                 allele so as to obtain a fingerprint of the
                 unknown allele; and

           (iii) comparing the fingerprint with fingerprints of
15   known alleles.

2.   A method according to claim 1 wherein the
polyallelic gene is a human leucocyte antigen (HLA) gene.

20   3.   A method according to claim 2 wherein the HLA
gene is an HLA class I gene or an HLA class II gene.

        4.   A method according to claim 1, 2 or 3 wherein the
panel of probes consists of from 20 to 70 probes.
25

        5.   A method according to claim 3 wherein the HLA
gene is an HLA class I gene and the panel of probes
comprises from 20 to 40 probes which each recognises the
motif recognised by one of the probes set out in Table 1.
30

        6.   A method according to claim 5 wherein each of the
probes has a sequence at least 40% identical to either (i)
a sequence of one of the probes set out in Table 1 or (ii)
a sequence complementary to a sequence of one of the probes
35 set out in Table 1.

        7.   A method according to claim 5 wherein the panel

of probes comprises the 40 probes set out in Table 1.

8.   A method according to any one of the preceding claims wherein the unknown allele has been separated from a mixture of alleles of the polyallelic gene by

(i)   amplifying the alleles in the mixture of alleles;

(ii) hybridising single strands of the amplified alleles with a complementary strand of a reference allele to form duplexes; and

(iii) separating the duplexes.

9.   A method according to claim 8 comprising

(i)   amplifying the mixture of alleles employing a pair of primers in which one of the primers carries a ligand, so as to produce an amplified mixture of double-stranded alleles in which one of the strands carries a ligand;

(ii) contacting the amplified mixture of double-stranded alleles with a receptor on a solid support under conditions such that the ligand binds to the receptor;

(iii) separating the mixture of double-stranded alleles into single-strands and removing the strands that are not bound to the support by the ligand;

(iv) recovering the remaining strands from the support;

(v)   mixing the recovered strands with a complementary strand of a reference allele so as to form duplexes; and

(vi) separating the duplexes.

10.   A method according to claim 9, modified by recovering the strands that do not bind to the support instead of those that bind to the support, and mixing these recovered strands with the reference allele strand in step
5   (v).

11.   A method according to any one of the preceding claims wherein the unknown allele is from a prospective donor or a prospective recipient in a tissue or organ
10   transplant operation.

12.   A method according to any one of the preceding claims wherein each probe in the panel recognises a sequence motif that is present in from 2 to 20 alleles of
15   the polyallelic gene.

13.   A kit for identifying an unknown allele of a polyallelic gene, which kit comprises a panel of probes, each of which probes recognises a sequence motif that is
20   present in some alleles of the polyallelic gene but not in others.

14.   A kit according to claim 13 also comprising a database which indicates which probes in the panel
25   recognise each allele of the polyallelic gene.

15.   A kit according to claim 13 or 14 also comprising a known allele as a control.

30   16.   A kit according to claim 13, 14 or 15 wherein the panel of probes is selected so as to recognise motifs of an HLA gene.

17.   A kit according to claim 16 wherein the HLA gene
35   is an HLA class I gene or an HLA class II gene.

18.   A kit according to any one of claims 13 to 17

wherein the panel consists of from 20 to 70 probes.

19.   A kit according to claim 17 wherein the HLA gene
is an HLA class I gene and the panel of probes comprises
5 from 20 to 40 probes which each recognises the motif
recognised by one of the probes set out in Table 1.

20.   A kit according to claim 19 wherein each of the
probes has a sequence at least 40% identical to either (i)
10 a sequence of one of the probes set out in Table 1 or (ii)
a sequence complementary to a sequence of one of the probes
set out in Table 1.

21.   A kit according to claim 19 wherein the panel of
15 probes comprises the 40 probes set out in Table 1.

22.   A kit according to any one of claims 13 to 21
wherein each probe in the panel recognises a sequence motif
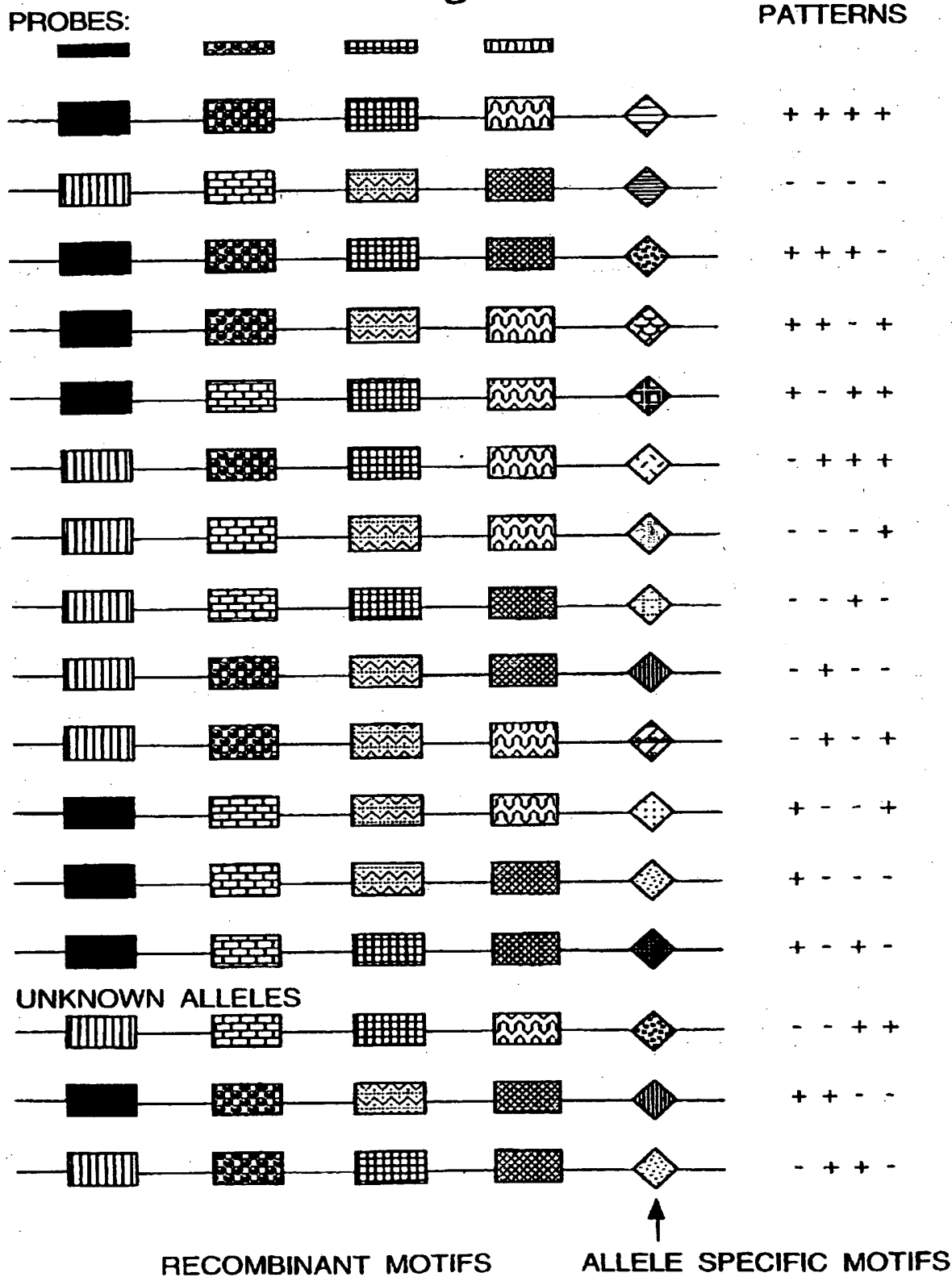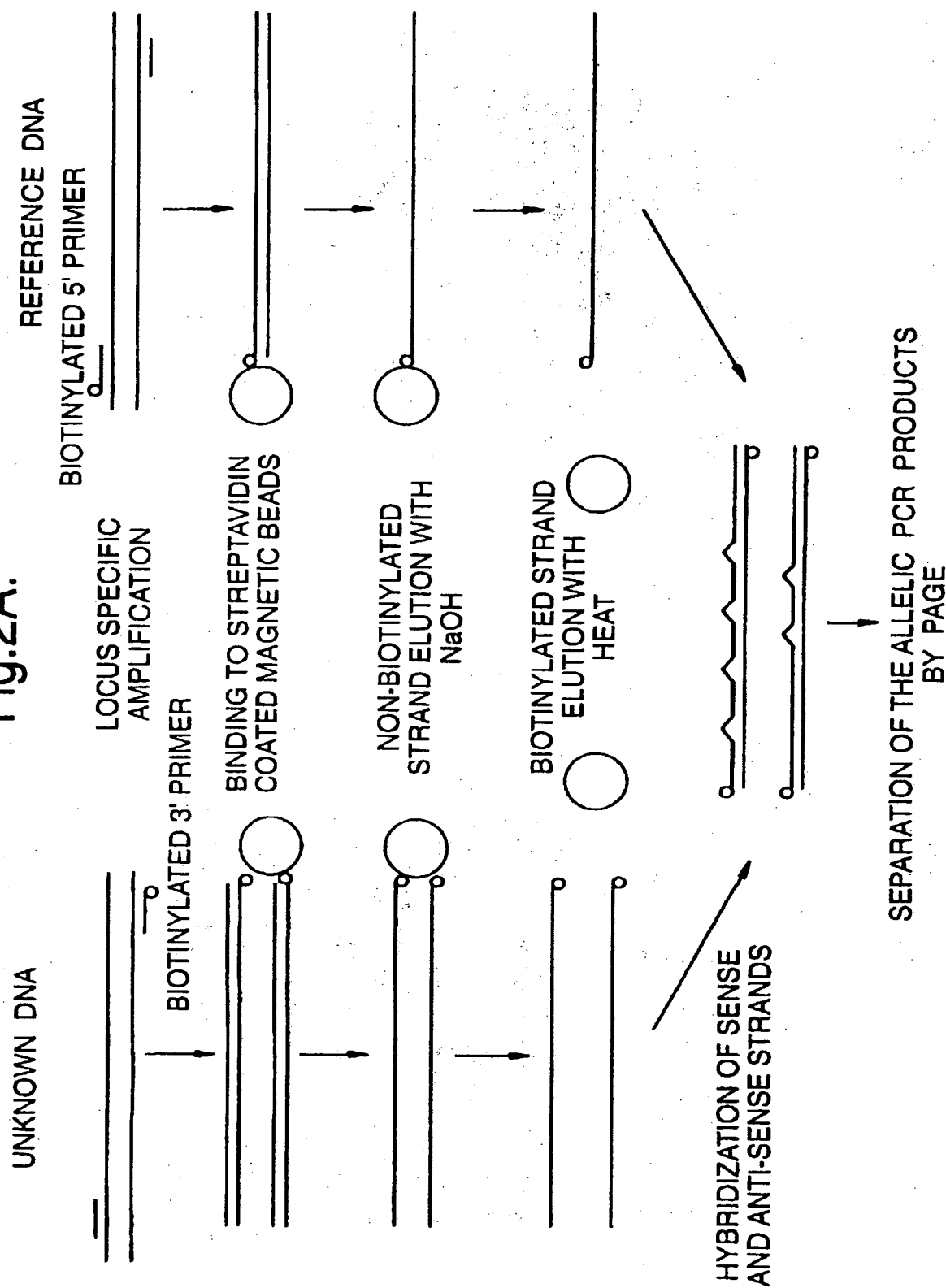that is present in from 2 to 20 alleles of the polyallelic
20 gene.

Fig.1.

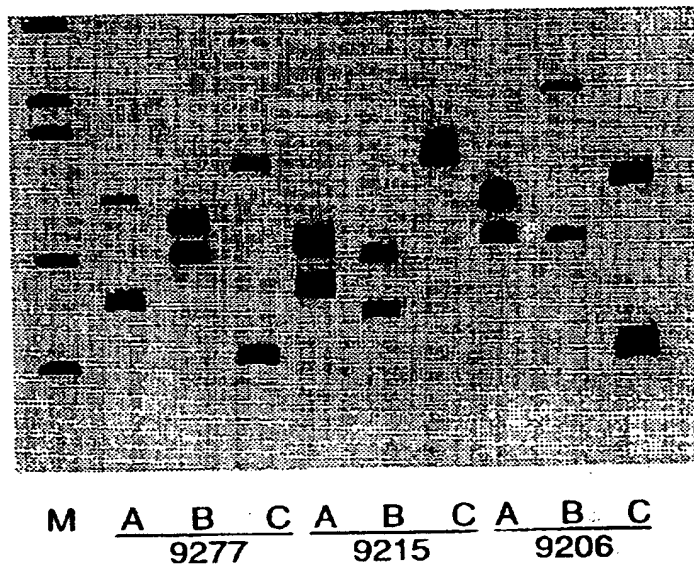PROBES: PATTERNS

+ + + +

- - - -

+ + + -

+ + - +

+ - + +

- + + +

- - - +

- - + -

- + - -

- + - +

+ - - +

+ - - -

+ - + -

UNKNOWN ALLELES

- - + +

+ + - -

- + + -

RECOMBINANT MOTIFS   ALLELE SPECIFIC MOTIFS

Fig.2A.

REFERENCE DNA

BIOTINYLATED 5' PRIMER

LOCUS SPECIFIC
AMPLIFICATION

BIOTINYLATED 3' PRIMER

BINDING TO STREPTAVIDIN
COATED MAGNETIC BEADS

NON-BIOTINYLATED
STRAND ELUTION WITH
NaOH

BIOTINYLATED STRAND
ELUTION WITH
HEAT

HYBRIDIZATION OF SENSE
AND ANTI-SENSE STRANDS

UNKNOWN DNA

SEPARATION OF THE ALLELIC PCR PRODUCTS
BY PAGE

Fig.2B.

M   A B C   A B C   A B C
   ‾‾‾‾‾‾‾‾‾   ‾‾‾‾‾‾‾‾‾   ‾‾‾‾‾‾‾‾‾
      9277   9215   9206

Fig.3.

Fig.4.